UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/814,382	03/31/2004	Gautam Das	307517.01	5509

22971   7590   10/11/2007

MICROSOFT CORPORATION
ONE MICROSOFT WAY
REDMOND, WA 98052-6399

EXAMINER
LOVEL, KIMBERLY M

ART UNIT	PAPER NUMBER
2167	

NOTIFICATION DATE	DELIVERY MODE
10/11/2007	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

roks@microsoft.com
ntovar@microsoft.com
a-rydore@microsoft.com

PTOL-90A (Rev. 04/07)

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE **3** MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on *25 June 2007*.

2a) ☒ This action is **FINAL**.   2b) ☐ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) *1-28 and 32* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) *1-28 and 32* is/are rejected.

7) ☐ Claim(s) _____ is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All   b) ☐ Some * c) ☐ None of:

      1. ☐ Certified copies of the priority documents have been received.

      2. ☐ Certified copies of the priority documents have been received in Application No. _____.

      3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

1.   This communication is responsive to the Amendment filed 25 June 2007.

2.   Claims 1-28 and 32 are pending in the current application. In the
Amendment filed 25 June 2007, claims 1-9, 11, 12, 15-18 and 20 are amended
and claims 29-31 and 33-34 are canceled. This action is made Final.

3.   The rejections of claims 1-28 and 32 as being anticipated by US Pat No
6,865,567 to Oommen are withdrawn as necessitated by applicants'
amendments.

Claim Objections

4.   Claims 9 and 17 are objected to because of the following informalities:

Claim 9 recites the term "contant." It seems as if the term should be
"constant."

Claim 17 recites the term "erros." It seems as if the term should be
"errors."

Appropriate correction is required.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35

U.S.C. 102 that form the basis for the rejections under this section made in this

Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

5.   **Claims 1-6 and 10 are rejected under 35 U.S.C. 102(b) as being**

anticipated by US Patent No 6,278,989 to Chaudhuri et al (hereafter

Chaudhuri).

Referring to claim 1, Chaudhuri discloses in database system, a

sampling method for constructing a data structure based on the contents of a

database comprising:

selecting an initial sample of data [initial sample of data values] from the

database, the initial sample of data including one or more subparts [bins of the

histogram] (see column 9, lines 66 – column 10, line 1);

cross-validating a plurality of subparts of the initial data sample, the cross-

validating associated with an error corresponding to a subpart [desired degree of

accuracy] (see column 9, line 59 – column 10, line 4);

sorting substantially simultaneously with the cross-validating the plurality

of subparts to generate a plurality of cross-validation errors (see column 11, lines

46-57);

generating an estimated block size based on the sorting and cross-

validating (see column 11, lines 14-34);

selecting an additional sample of data, wherein the size of the selected additional sample of data corresponds to the generated estimated block size see column 11, lines 14-34;

merging the additional sample of data with the initial sample of data (see column 11, lines 35-38).

Referring to claim 2, Chaudhuri discloses the method of claim 1 wherein the cross-validating includes cross-validating subparts of data that are of different sizes (see column 9, line 59 – column 10, line 13).

Referring to claim 3, Chaudhuri discloses the method of claim 1 wherein the cross-validating and sorting are combined in a single step (see column 11, lines 47-57).

Referring to claim 4, Chaudhuri discloses the method of claim 3 wherein the single step includes:

dividing the initial sample of data into multiple subparts [bins of the histogram]; sorting and cross-validating the multiple subparts recursively (see column 9, line 59 – column 10, line 4).

Referring to claim 5, Chaudhuri discloses the method of claim 4 wherein the single step further includes:

building a histogram [equi-height k-histogram] for at least a first subpart and a second subpart of the initial sample of data; testing the histogram of the first subpart against the second subpart to generate a cross-validation error estimate for a sample size corresponding to the initial sample of data (see column 9, line 66 – column 10, line 5).

Referring to claim 6, Chaudhuri discloses the method of claim 5 further comprising reusing parts of the initial sample of data to generate different cross-validation error estimates, each of the cross-validation error estimates corresponding to an associated sample size (see column 9, line 66 – column 10, line 5).

Referring to claim 10, Chaudhuri discloses a computer readable medium for performing computer instructions to implement the method of claim 1 (see column 1, lines 55-59).

Claim Rejections - 35 USC § 103

6.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7.   **Claims 11-28 and 32 are rejected under 35 U.S.C. 103(a) as being unpatentable over US Pat No 6,865,567 to Oommen et al (hereafter Oommen) in view of US Patent No 6,278,989 to Chaudhuri et al.**

Referring to claim 11, Oommen discloses a database system for constructing histograms based on sampling the contents of the database comprising:

a) a database management component that gathers block size data segments from the database which in aggregate form a first sample of data having a first size [first phase – x number of tuples] (see column 21, lines 16-19);

b) a histogram construction component that forms a first histogram from the first sample of data (see Fig 11); and

c) a correlation component (see column 21, lines 19-40);

wherein said database management component gathers an additional sample of data used by said histogram construction component in creating a resultant histogram corresponding to a combination of the additional sample and the initial sample of data, the size of the additional sample being based on the cross-validation errors (see column 22, line 48 – column 23, line 17).

However, Oommen fails to explicitly disclose the further limitation wherein the correlation component cross-validates a plurality of subparts of the initial sample of data and sorts substantially simultaneously with the cross-validating the plurality of subparts to generate a plurality of cross-validation errors. Chaudhuri discloses using adaptive random sampling with cross-validation to determine when enough data of a database has been sampled to construct histograms on one or more columns of one or more tables of the database within a desired or predetermined degree of accuracy (see abstract), including the further limitation of wherein the correlation component cross-validates a plurality of subparts of the initial sample of data and sorts substantially simultaneously with the cross-validating the plurality of subparts to generate a plurality of cross-validation errors (see column 11, lines 46-57).

It would have been obvious to one of ordinary skill in the art at the time of the invention to utilize the concept of performing sorting and cross-validation simultaneously as disclosed by Chaudhuri with the sorting and cross-validation steps of Oommen. One would have been motivated to do so in order to increase the efficiency of collecting database samples.

Referring to claim 12, Oommen/Chaudhuri discloses the system of claim 11 wherein the resultant histogram is formed by the histogram construction component based on data gathered in the first sample of data and the additional sample of data (Oommen: see column 22, lines 60-67).

Referring to claim 13, Oommen/Chaudhuri discloses the system of claim 11 wherein the first sample of data and the additional sample of data are randomly retrieved block samples (Oommen: see column 20, lines 60-67).

Referring to claim 14, Oommen/Chaudhuri discloses the system of claim 11 wherein histogram construction component sorts the data in said first sample of data as it constructs the first histogram (Oommen: see column 22, lines 60-67 and Fig 11).

Referring to claim 15, Oommen/Chaudhuri discloses the system of claim 11 wherein the correlation component determines the cross-validation errors by cross correlating the contents of the first histogram with other data in said first sample of data to determine an initial sufficiency (Oommen: see Fig 11 and Fig 19).

Referring to claim 16, Oommen/Chaudhuri discloses the system of claim 15 wherein the first sample of data is sub-divided to form the subparts used to

form histograms of differing sizes that are cross correlated to find a cross-validation error relating to said differing sample sizes (Oommen: see Fig 19).

Referring to claim 17, Oommen/Chaudhuri n discloses the system of claim 15 wherein the first sample of data is sub-divided to form additional subparts of smaller size that are used to form other histograms that are cross correlated for use in finding cross-validation errors relating to sample sizes for use in determining a size of the additional sample of data to gather from the database (Oommen: see Fig 11).

Referring to claim 18, Oommen discloses in a database system, a sampling method for constructing a histogram based on the contents of a database comprising:

a) gathering an initial sample [first phase with x number of tuples] (see column 21, lines 16-19 and column 22, lines 60-67) of data from the database and creating a histogram from said initial sample;

b) gathering a second sample of data from the database for comparison with said first histogram [second phase] (see column 21, lines 19-40);

c) determining an initial sufficiency of the data gathered from the database that is based on a comparison of the second sample with the first histogram (see column 21, lines 19-40); and

d) if the determination of initial sufficiency indicates the data in said initial and second samples is adequate to represent the database, combining the initial and second samples to form a resultant histogram, but if the determination of initial sufficiency indicates the initial and second samples are inadequate to

represent the database, gathering an additional data sample to combine with the
initial and second samples to form the resultant histogram wherein a size of the
additional data sample is based on the initial sufficiency determination (see
column 22, line 48 – column 23, line 17).

However, Oommen fails to explicitly disclose the further limitation wherein
the correlation component cross-validates a plurality of subparts of the initial
sample of data and sorts substantially simultaneously with the cross-validating
the plurality of subparts to generate a plurality of cross-validation errors.
Chaudhuri discloses using adaptive random sampling with cross-validation to
determine when enough data of a database has been sampled to construct
histograms on one or more columns of one or more tables of the database within
a desired or predetermined degree of accuracy (see abstract), including the
further limitation of wherein the correlation component cross-validates a plurality
of subparts of the initial sample of data and sorts substantially simultaneously
with the cross-validating the plurality of subparts to generate a plurality of cross-
validation errors (see column 11, lines 46-57).

It would have been obvious to one of ordinary skill in the art at the time of
the invention to utilize the concept of performing sorting and cross-validation
simultaneously as disclosed by Chaudhuri with the sorting and cross-validation
steps of Oommen. One would have been motivated to do so in order to increase
the efficiency of collecting database samples.

Referring to claim 19, Oommen/Chaudhuri discloses the method of claim 18 wherein the data is gathered in blocks from random storage locations within the database (Oommen: see column 20, lines 60-67).

Referring to claim 20, Oommen discloses in a database system, a system for constructing a data structure based on the contents of a database comprising:

a) means for gathering an initial sample [first phase] of data from the database and creating a first data structure [histogram] from said initial sample (see column 21, lines 16-19 and column 22, lines 60-67);

b) means for determining an initial sufficiency of the data gathered from the database that is based on a comparison of the first data structure and other data in the initial sample not used to create the first data structure (see column 21, lines 19-40); and

c) means for forming a resultant data structure by gathering an additional sample of data from the database and using the additional amount of data to form the resultant data structure wherein the amount of data gathered in the additional sample is based on the initial sufficiency determination (see column 22, line 48 – column 23, line 17).

However, Oommen fails to explicitly disclose the further limitation wherein the correlation component cross-validates a plurality of subparts of the initial sample of data and sorts substantially simultaneously with the cross-validating the plurality of subparts to generate a plurality of cross-validation errors. Chaudhuri discloses using adaptive random sampling with cross-validation to

determine when enough data of a database has been sampled to construct

histograms on one or more columns of one or more tables of the database within

a desired or predetermined degree of accuracy (see abstract), including the

further limitation of wherein the correlation component cross-validates a plurality

of subparts of the initial sample of data and sorts substantially simultaneously

with the cross-validating the plurality of subparts to generate a plurality of cross-

validation errors (see column 11, lines 46-57).

It would have been obvious to one of ordinary skill in the art at the time of

the invention to utilize the concept of performing sorting and cross-validation

simultaneously as disclosed by Chaudhuri with the sorting and cross-validation

steps of Oommen. One would have been motivated to do so in order to increase

the efficiency of collecting database samples.

Referring to claim 21, Oommen/Chaudhuri discloses the system of claim

20 wherein the resultant data structure is formed based on data gathered in the

initial sample and the additional sample (Oommen: see column 21, lines 19-40

and column 23, line 17).

Referring to claim 22, Oommen/Chaudhuri discloses the system of claim

21 wherein the first and resultant data structures are histograms (Oommen: see

column 22, lines 60-67).

Referring to claim 23, Oommen/Chaudhuri discloses the system of claim

20 wherein the initial data sample is made up of randomly retrieved block

samples that form a first amount of data that is divided in half to provide data to

form the data structure and data to cross correlate against the first data structure

(Oommen: see column 20, lines 60-67).

Referring to claim 24, Oommen/Chaudhuri discloses the system of claim

23 wherein the initial data samples is sorted and used to form two histograms

(Oommen: see Fig 11).

Referring to claim 25, Oommen/Chaudhuri discloses the system of claim

24 wherein an error metric of the two histograms are formed by cross correlating

the contents of the two histograms to determine the initial sufficiency (Oommen:

see Fig 19).

Referring to claim 26, Oommen/Chaudhuri discloses the system of claim

25 wherein the initial data sample is further sub-divided to form sub-samples

used to form other histograms of differing sample sizes that are cross correlated

to find an error metric relating to said differing sample sizes (Oommen: see Fig

19).

Referring to claim 27, Oommen/Chaudhuri discloses the system of claim

26 wherein the initial and second data samples are further sub-divided to form

additional sub-samples of smaller size that are used to form other histograms

that are cross correlated for use in finding an error metric relating to sample sizes

for use in determining a size of the additional sample of data to gather from the

database (Oommen: see Fig 11).

Referring to claim 28, Oommen/Chaudhuri discloses the system of claim

24 additionally comprising means for estimating distinct values of an attribute of

the initial and second samples by eliminating records from the blocks that are

duplicated within a given block and estimating distinct values by categorizing

attributes as rarely or frequently occurring within the database (Oommen: see

column 7, lines 40-49).

Referring to claim 32, Oommen/Chaudhuri discloses a computer

readable medium for performing computer instructions to implement the method

of claim 20 (Oommen: see column 117, lines 12-24).

8.   **Claims 7-9 are rejected under 35 U.S.C. 103(a) as being unpatentable
over US Patent No 6,278,989 to Chaudhuri et al as applied to claim 6 above,
and further in view of US Pat No 6,865,567 to Oommen et al.**

Referring to claim 7, Chaudhuri discloses computing means of the

different cross-validation error estimates for each of the associated sample sizes

(see column 9, line 59 – column 10, line 14). Chaudhuri fails to explicitly disclose

the further limitations of determining a best fit of the means of the different cross-

validation error estimates and estimating the block size based on the determined

best fit. Oommen discloses steps for creating histograms (see abstract),

including the further limitations of determining a best fit of the means of the

different cross-validation error estimates; and estimating the block size based on

the determined best fit (see column 33, line 31 – column 34, line 16).

It would have been obvious to one of ordinary skill in the art at the time of

the invention to utilize the determining a best fit of the means as disclosed by

Oommen with the means of Chaudhuri. One would have been motivated to do

so in order to increase the efficiency of collecting database samples.

Referring to claim 8, the combination of Chaudhuri and Oommen (hereafter Chaudhuri/Oommen) discloses the method of claim 7 wherein determining the best fit includes identifying a best fitting curve associated with the means of the different cross-validation error estimates (Chaudhuri: see column 8, lines 1-9).

Referring to claim 8, Chaudhuri/Oommen discloses the method of claim 4-8 wherein identifying a best fitting curve includes: generating the best fitting curve of the form $\Delta 2 = c/r$, wherein c is a constant, $\Delta 2$ is an average squared cross-validation error observed for a given sample size, and r represents the given sample size; estimating the block size based on the constant c (Chaudhuri: see column 8, lines 1-9).

Conclusion

9.   **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of

time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire

THREE MONTHS from the mailing date of this action. In the event a first reply is

filed within TWO MONTHS of the mailing date of this final action and the advisory

action is not mailed until after the end of the THREE-MONTH shortened statutory

period, then the shortened statutory period will expire on the date the advisory

action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be

calculated from the mailing date of the advisory action. In no event, however, will

the statutory period for reply expire later than SIX MONTHS from the mailing

date of this final action.

Contact Information

Any inquiry concerning this communication or earlier communications from

the examiner should be directed to Kimberly Lovel whose telephone number is

(571) 272-2750. The examiner can normally be reached on 8:00 - 4:00.

If attempts to reach the examiner by telephone are unsuccessful, the

examiner's supervisor, John Cottingham can be reached on (571) 272-7079.

The fax phone number for the organization where this application or proceeding

is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from

the Patent Application Information Retrieval (PAIR) system. Status information

for published applications may be obtained from either Private PAIR or Public

PAIR. Status information for unpublished applications is available through

Private PAIR only. For more information about the PAIR system, see http://pair-

direct.uspto.gov. Should you have questions on access to the Private PAIR

system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-

free). If you would like assistance from a USPTO Customer Service

Representative or access to the automated information system, call 800-786-

9199 (IN USA OR CANADA) or 571-272-1000.

Kimberly Lovel
Examiner
Art Unit 2167

1 October 2007
kml

JOHN COTTINGHAM
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100